

**[0030]** The front-end server **102** receives the personalized search results and the personalized advertisements, and combines them (or a subset of each) to form a web page (results page) having some number of the documents from the search results and some number of the advertisements. This results page is returned to the client **118**, where its rendered and displayed to the user, typically in the window of a browser or similar application (depending on client device). The personalized advertisements can be displayed next to the search result lists in a side panel, in a separate frame of the window, or in any other graphical format deemed appropriate.

**[0031]** The next sections describe the construction and use of user profiles to personalize search results, and the construction and use of the search profiles to personalize advertisement.

#### **[0032]** Creation and Maintenance of User Profiles

**[0033]** A user profile describes the user's interests in a manner that can be used to personalize the results of any particular search query. The user profile can be derived from information that is explicitly provide by the user (e.g., designation of interests or topics in a directory), or information that is inferred from the user's behaviors and interactions with the search engine **104**, or information that is inferred from the user's online relationships (e.g., websites or pages associated with the user's IP address).

**[0034]** With respect to information derived from the user's interaction with the search engine **104**, prior search activities (both search queries themselves, and user access or non-access to the results) provide useful hints about the user's interests. **FIG. 2** provides an overview of various sources of information that are beneficial for user profile construction. For example, previously submitted search queries **201** are very helpful in profiling a user's interests. If a user has submitted multiple search queries related to diabetes, it is more likely than not that this is a topic of interest to the user. If the user subsequently submits a query including the term "organic food", it can be reasonably inferred that he may be more interested in those organic foods that are helpful in fighting diabetes. Similarly, the universal resource locators (URL) **203** associated with the search results in response to the previous search queries and their corresponding anchor texts **205**, especially for search result items that have been selected or "visited" by the user (e.g., downloaded or otherwise viewed by the user), are helpful in determining the user's preferences. When a first page contains a link to a second page, and the link has text associated with it (e.g., text neighboring the link), the text associated with the link is called "anchor text" with respect to the second page. Anchor text establishes a relationship between the text associated with a URL link in a document and another document to which the URL link points. The advantages of anchor text include that it often provides an accurate description of the document to which the URL link points, and it can be used to index documents that cannot be indexed by a text-based search engine, such as images or databases. In addition, a count may be maintained for each URL that is associated with the user's search results, and URLs receiving high counts are identified or otherwise analyzed in the user profile.

**[0035]** After receiving search results, the user may click on some of the URL links, thereby downloading the docu-

ments referenced by those links, so as to learn more details about those documents. Certain types of general information **207** can be associated with a set of user selected or use identified documents. For purposes of forming a user profile, the identified documents from which information is derived for inclusion in the user profile may include: documents identified by search results from the search engine, documents accessed (e.g., viewed or downloaded, for example using a browser application) by the user (including documents not identified in prior search results), documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the user, or any subset of such documents.

**[0036]** The general information **207** about the identified documents is also useful information about the user's preferences and interests. General information includes information such as the document format of accessed documents (e.g., HTML, plain text, portable document format (PDF), Microsoft Word), date information, creator information, and other metadata.

**[0037]** Activity information **209** describes the user's activities with respect to the user selected documents (sometimes herein called the identified documents). This information describes factors such as how long the user spent viewing the document, the amount of scrolling activity on the document, and whether the user has printed, saved or bookmarked the document, and thus also suggests the importance of the document to the user as well as the user's preferences. In some embodiments, information about user activities **209** is used when weighting the importance of information extracted or derived from the user identified documents. In some embodiments, information about user activities **209** is used to determine which of the user identified documents to use as the basis for deriving the user profile. For example, information **209** may be used to select only documents that received significant user activity (in accordance with predefined criteria) for generating the user profile, or information **209** may be used to exclude from the profiling process documents that the user viewed for less than a predefined threshold amount of time.

**[0038]** The content of identified documents from previous search activities is a rich source of information about a user's interests and preferences. Key terms appearing in the identified documents and their frequencies with which they appear in the identified documents are not only useful for indexing the document, but are also a strong indication of the user's personal interests, especially when they are reinforce other types of user information discussed above. In one embodiment, instead of the whole documents, sampled content **211** from the identified documents is extracted for the purpose of user profile construction, to save storage space and computational cost. In another embodiment, various information related to the identified documents may be classified to constitute category information **213** about the identified documents. More discussion about content sampling, the process of identifying key terms in an identified document and the usage of the category information is provided below.

**[0039]** Optionally, a user may choose to offer personal information **215**, including demographic and geographic information associated with the user, such as the user's age or age range, educational level or range, income level or